

# Discussion Papers

# 831

Christian von Hirschhausen • Astrid Cullmann

**Next Stop: Restructuring?  
A Nonparametric Efficiency Analysis of  
German Public Transport Companies**

Berlin, October 2008

Opinions expressed in this paper are those of the author and do not necessarily reflect views of the institute.

## IMPRESSUM

© DIW Berlin, 2008

DIW Berlin  
German Institute for Economic Research  
Mohrenstr. 58  
10117 Berlin  
Tel. +49 (30) 897 89-0  
Fax +49 (30) 897 89-200  
<http://www.diw.de>

ISSN print edition 1433-0210  
ISSN electronic edition 1619-4535

Available for free downloading from the DIW Berlin website.

Discussion Papers of DIW Berlin are indexed in RePEc and SSRN.  
Papers can be downloaded free of charge from the following websites:

[http://www.diw.de/english/products/publications/discussion\\_papers/27539.html](http://www.diw.de/english/products/publications/discussion_papers/27539.html)  
<http://ideas.repec.org/s/diw/diwwpp.html>  
[http://papers.ssrn.com/sol3/JELJOUR\\_Results.cfm?form\\_name=journalbrowse&journal\\_id=1079991](http://papers.ssrn.com/sol3/JELJOUR_Results.cfm?form_name=journalbrowse&journal_id=1079991)

# **Next Stop: Restructuring? A Nonparametric Efficiency Analysis of German Public Transport Companies**

**Christian von Hirschhausen<sup>1</sup> and Astrid Cullmann<sup>2</sup>**

October 2008

## **Abstract**

In this paper, we present a nonparametric comparative efficiency analysis of 179 communal public transport bus companies in Germany (1990-2004). We apply both deterministic data envelopment analysis (DEA) and bootstrapping to test the robustness of our estimates and to test the hypothesis of global and individual constant returns to scale. We find that the average technical efficiency of German bus companies is relatively low. We observe that the industry appears to be characterized by increasing returns to scale for smaller companies. These results would imply increasing pressure on bus companies to restructure.

Keywords: public transport, buses, efficiency analysis, nonparametric methods, DEA, bootstrapping

JEL-codes: L11, L51, L92, C14

---

<sup>1</sup> Dept. of International Economics, DIW Berlin (German Institute for Economic Research)  
Mohrenstraße 58, 10117 Berlin (Germany), [chirschhausen@diw.de](mailto:chirschhausen@diw.de), tel.: +49-30-89789-343 (-341)  
fax.: +49-30-89789-108 (-200),  
Dresden University of Technology, Chair of Energy Economics and Public Sector Management  
D-01062 Dresden (Germany)

<sup>2</sup> Dept. of International Economics, DIW Berlin (German Institute for Economic Research)  
Mohrenstraße 58, 10117 Berlin (Germany), [acullmann@diw.de](mailto:acullmann@diw.de), tel.: +49-30-89789-679



# **1 Introduction**

European policy makers face increasing pressure to reform both regional and national public transport systems. In Germany, structural reform is particularly urgent, since financial pressure on this sector, which traditionally requires subsidies, is mounting, and competitive forces are ever more active. At the European level where general policy guidelines on state aid and other instruments are formulated, an important issue in public transport economics is whether the sector is characterized by increasing, decreasing, or constant returns to scale. A survey of the extant literature appears to reach the conclusion that within smaller and medium-sized bus companies, increasing returns to scale prevail, whereas the assessment on larger and very large firms is more uncertain. Thus, Berechman (1993) finds that smaller bus companies (those with as many as 200 buses) are characterized by increasing returns to scale, but that the opposite may be the case for very large companies. For example, Chicago (2,500 buses) and New York City's MTA (3,000 buses) can be characterized by decreasing returns to scale.

Increasing returns to scale for smaller companies are confirmed by Viton (1981), Cowie and Asenova (1999), Filippini and Prioni (2003) and Farsi et al. (2006), among others. For surveys on the topic see Berechman (1993), De Borger et al. (2002) and Piacenza (2001). Increasing returns to scale would suggest that smaller and medium-sized companies can increase their efficiency by growing, merging with other companies, or achieving the synergy effects in other ways. In fact, a concentration of smaller firms towards larger units can be observed in countries that have liberalized their public bus sectors, such as the UK (Cowie, 2002). Graham et al. (2003) who examined economies of scale and density in 17 urban rail transport systems around the world suggest constant returns to scale but increasing returns to density.

The German public transport system is traditionally small-scale, with about 1,000 mainly communal service providers. Contrary to other countries, efficiency benchmarking has not been intensively carried out for the sector at large.<sup>1</sup> There is one study conducted by Hanusch and Canter (1991) analyzing the performance of German bus companies within a multiple-sector analysis. It is generally considered that a lack of transparency and an information asymmetry exist between the principal (public policy maker, “Aufgabenträger”) and the agent (public transport company). Public transport companies argue about specific institutional, economic and structural factors that justify a high level of subsidies. Cost-based compensation instead of incentive-based mechanisms still predominates. Traditionally, public bus transport was provided at the communal level, and rail transport was formerly the responsibility of the Federal State Railway (responsibility for public local rail transport was transferred to the 16 Federal States in 1994 under “Regionalization”). Public transport has a low cost coverage (estimated around 40%), with the bus system faring somewhat better than the rail system. Both bus and rail transport are highly regulated; in particular, bus licenses for regional transport services are difficult to obtain. As more local concessions are tendered, new entrants or larger bus companies obtain higher market shares. Communal bus companies also try to cooperate across borders, e.g. to obtain economies in third-party procurement.

Regarding the incentive structures, it is fair to say that - as in other countries - the industry is unionized, which is also due to the dominant communal ownership structure. Yet there has been a general tendency to reduce employment over the period under consideration. Bus companies are „agents“ and act on behalf of a “principal”, regional or city administrations („Aufgabenträger“) who order a certain level of service. Changing network length, therefore, is not in the simple discretionary decision power of the bus companies.

This paper provides a preliminary nonparametric efficiency benchmarking of public transport in Germany between 1990 and 2004 to an unbalanced sample of 179 medium and larger bus

---

<sup>1</sup> Business studies on efficiency have been carried out over the last 20 years by Helmut Leuthardt, and published in “Der Nahverkehr” (see e.g. Leuthardt 1986, 2005).

transportation companies. We apply recent theoretical developments in statistical inference for nonparametric efficiency estimation. In empirical efficiency analysis the tradeoff always exists between the restrictive (but when consistent, more efficient) parametric and the more robust (but inefficient) nonparametric approaches. There exists a wide range of literature comparing both approaches. Thus, Lovell (1993) provides a detailed introduction. Ferrier et al. (1990) e.g. assess the strength and weakness of both approaches by means of an empirical cost efficiency analysis in banking. Bjurek et al. (1990) compare both approaches within the framework of service production. A more recent example is Cullinane et al. (2006) who provide a technical efficiency analysis of container ports comparing the parametric stochastic frontier analysis (SFA) and DEA as a nonparametric approach, pointing out the strengths and weaknesses associated with each approach.

The goal of this paper is to provide an objective assessment of the comparative efficiency scores of public transport companies, thus decreasing the information asymmetry, and to contribute constructive input to the debate on public transport reforms. The paper is structured as follows: the next section describes the methodology, focusing on the latest developments of nonparametric estimation. The robustness of the results is analyzed by means of bootstrapping algorithms, and newly-developed returns to scale tests<sup>2</sup>. Section 3 introduces the data and the concrete model specifications, with a focus on supply-oriented models using seat and bus kilometers as different output variables. Most of the results confirm our initial hypothesis of increasing scale economies for small and medium-sized bus companies (Section 4). Section 5 concludes.

---

<sup>2</sup> A similar study is carried out for Canadian urban transit systems by Boame (2004) who uses a bootstrap data envelopment analysis for the period 1990-1998. He found that most Canadian transit systems experience increasing returns to scale.

## 2 Methods

### 2.1 Data Envelopment Analysis (DEA)

To measure the relative efficiency of the German public transit bus companies, we apply nonparametric techniques that have proven useful in a number of other sectors and applications.<sup>3</sup> Applied empirical work on efficiency and productivity measurement of individual firms is always confronted with the sensitivity of the results to the different approaches and assumptions. Therefore, to present the most robust image, we apply different nonparametric model specifications and, in a second step, test our empirical results using recent developments and approaches in statistical inference for nonparametric frontiers (Simar and Wilson, 2000, 2002, 2007). This nonparametric approach of efficiency measurement of different decision-making units (DMUs) relies on a production frontier which is defined as the geometrical locus of optimal production plans (see Simar and Wilson, 1998). The individual efficiencies of the firms relative to this production frontier are calculated by means of distance functions. The input distance function  $d_i$  is defined on the input set  $L(y)$  as

$$d_i(x, y) = \max\{\rho : (x / \rho) \in L(y)\} \quad (1)$$

where  $\rho$  is the scalar distance and considers by how much the input vector may be proportionally contracted with the output vector held fixed (see Coelli, 2000) remaining within the feasible input set;  $d_i(x, y)$  will take a value which is greater than or equal to one if the input vector  $x$  is an element of the feasible input set  $L(y)$ , representing the set of all input vectors. In addition,  $d_i(x, y) = 1$  if  $x$  is located on the inner boundary of the input set. The input oriented measure of technical efficiency can be expressed by  $TE = 1 / d_i(x, q)$ . Färe

---

<sup>3</sup> For a survey on the theoretical literature see e.g. Cooper et al. (2004).



et al. (1985) used linear programming methods to construct nonparametric distance functions for the measurement of technical efficiency (see Coelli, 2000).

The two common nonparametric envelopment methods are DEA and the free disposal hull (FDH) which was proposed by Deprins et al. (1984). This paper employs DEA that involves the use of linear programming methods to construct a nonparametric piecewise linear surface or frontier over the data and measures the efficiency for a given unit relative to the boundary of the convex hull of the input output vectors. Coelli et al. (2005) introduces the DEA in an intuitive way using the ratio form (see Coelli et al., 2005 for a derivation). Using duality in linear programming the determination of the efficiency score of the  $i$ -th firm in a sample of  $N$  firms in the CRS model is equivalent to the following optimization:

$$\begin{aligned}
 & \min \theta_{\theta, \lambda} \\
 & s.t \\
 & -y_i + Y\lambda \geq 0 \\
 & \theta x_i - X\lambda \geq 0 \\
 & \lambda \geq 0
 \end{aligned} \tag{1}$$

where  $\lambda$  is a  $N \times 1$  vector of constants and  $X, Y$  represent input and output matrices respectively.  $\theta$  measures the radial distance between the observation  $x, y$  and the point on the frontier characterized by the level of inputs that should be reached to be efficient; thus is the efficiency score and is equal to  $TE = \theta = 1/d_i$ .  $\lambda$  determines the weights for the firms' inputs and outputs. The value  $\theta = 1$  ( $d_i = 1$ ) indicates that a firm is fully efficient and thus is located on the efficiency frontier. To determine efficiency measures under the assumption of variable returns to scale (VRS), a further convexity constraint  $\sum \lambda = 1$  must be considered.

We focus on the firms' technology and production processes to assess technical efficiency.<sup>4</sup> Calculations can be made using either an input-orientation or an output-orientation. Since it is reasonable to assume a fixed, exogenous output for public transport companies that have a legal duty to serve and supply certain areas with predefined intensity, we apply input-orientation (see Introduction for more details about the institutional environment justifying the assumption). It is also important for us to discover whether the technology exhibits constant, increasing or decreasing returns to scale. Thus we include technical inefficiency as well as scale inefficiency under the CRS approach. We note that by using the variable returns to scale (VRS) approach, the performance only reflects the pure technical efficiency of the decision-making units. In any case, finding a significant difference as a result of using either approach will indicate scale inefficiency. In our empirical analysis we start with a descriptive analysis looking at the difference and then test the nature of the technology of the communal bus companies by bootstrapping methods (see Section 2.3).

## **2.2 Non-discretionarity**

The traditional DEA approach can be modified to “capture” different production characteristics. We use in our empirical application the approach of non-discretionarity (see Banker and Morey, 1986). Within this framework we account for input variables that are not under the control of the bus companies in the short run, such as the characteristics of the service area. We adapt the DEA approach to find only radial reduction in the inputs over which the manager has discretionary control (see Coelli et al., 2005). We can rewrite Equation 1 for the CRS case as:

---

<sup>4</sup> If price data is available and one assumes a behavioral objective, such as cost minimization, it is possible to consider allocative efficiency and relate it to technical efficiency to measure the overall efficiency of the firms (see Coelli et al., 2005).

$$\begin{aligned}
& \min \theta_{\theta, \lambda} \\
& s.t \\
& -y_i + Y\lambda \geq 0 \\
& \theta x_i^D - X^D \lambda \geq 0 \\
& x_i^{ND} - X^{ND} \lambda \geq 0 \\
& \lambda \geq 0
\end{aligned} \tag{2}$$

where  $x_i^D$  and  $x_i^{ND}$  denote discretionary and non-discretionary sets respectively.  $\theta$  is now associated only with the discretionary inputs (i.e. seeking radial reductions in this subset).

### 2.3 Statistical tests by means of nonparametric efficiency scores

While the nonparametric deterministic envelopment estimators have been widely used to measure the relative efficiency of firms, DEA techniques have been also criticized for being deterministic and non-statistical. Consequently, they are sensitive to extreme values and outliers and cannot account for noise in the data. Other criticisms concern the robustness and validation of results. Today, sensitivity analysis and statistical inference based on the DEA estimator are available either by using asymptotic results or by means of simulation methods - the bootstrap approach (for a survey, see Simar and Wilson, 2000, 2007).<sup>5</sup> To resolve the issue of robustness, we apply the bootstrap algorithm established in Simar and Wilson (1998) first for the bias correction and the creation of confidence intervals and then for the test on returns to scale. Knowledge on the economies of scale and economies of density are important for policymakers and for business managers in the companies. It is therefore crucial to determine whether the underlying technology exhibits increasing, constant, or decreasing returns to scale. For the interpretation of results, information about the efficiency and the consistency of estimators is important. When taking a long-term view on the flexibility of input combinations, i.e. the CRS-assumption, there is a risk to obtain inconsistent efficiency

---

<sup>5</sup> In addition the issue was addressed by ex ante descriptive statistics, e.g. Wilson (1993). Cazals et al. (2002), explicitly introduced stochasticity, building a nonparametric estimator of the efficiency frontier, the order-m estimator, which is more robust to extreme values and outliers or noise in the data (see Section 3.1 for more details on outlier detection).

estimates because the underlying technology may in fact display variable returns to scale. On the other hand, there may be a loss of statistical efficiency if we assume variable returns to scale when in reality the technology exhibits global constant returns to scale (Simar and Wilson, 2002).<sup>6</sup>

In general returns to scale are measured by means of the distance functions and efficiency scores outlined in Section 2.1. One computes the distance estimator from the observed data for CRS, VRS and non increasing returns to scale (NIRS).<sup>7</sup> We choose to compare in a descriptive framework the different estimations under both assumptions (VRS and CRS) (see Färe et al., 1985) using this approach (see Section 4.1). However, one problem is that conclusions are drawn based on the estimated technology and not on the true technology.<sup>8</sup> Instead, Simar and Wilson (2002) have proposed to start with:

Test 1  $\rightarrow H_0$ : Production frontier is globally CRS vs.  $H_1$ : The production frontier is VRS.

If we cannot reject  $H_0$ , we may choose to accept the null hypothesis of CRS. If  $H_0$  is rejected we might perform another test before accepting  $H_1$ .

Test 2  $\rightarrow H_0$ : Production frontier is globally NIRS vs.  $H_1$ : The production frontier is VRS.

If we reject both  $H_0$ , the technology represents VRS.

---

<sup>6</sup> One approach is suggested by Färe et al. (1985) who verified local returns to scale by comparing the empirical efficiency scores estimated under different assumptions. Their approach has been criticized because of its failure to provide a formal statistical test of returns to scale (for a discussion of other approaches see Simar and Wilson, 2002).

<sup>7</sup> To determine efficiency measures under the assumption of non increasing returns to scale (NIRS), a further convexity constraint  $\sum \lambda \leq 1$  must be considered in the linear optimization (see Equation 1).

<sup>8</sup> More precisely, if we find for one particular observation that the ratio of the estimated CRS and the VRS score are smaller than one, then without a formal statistical testing procedure, it is not possible to determine whether this is due to non-constant returns to scale or due to sampling variation (see Simar and Wilson, 2002).

Our test statistic is the usual measure of scale efficiency (see Färe et al., 1985), which is the estimated ratio between the CRS and VRS efficiency scores (the inverse of the distance function  $1/d_i = \theta = TE$ ).<sup>9</sup>

$$\hat{\omega} = \frac{\hat{\theta}_n^{CRS}(x, y)}{\hat{\theta}_n^{VRS}(x, y)}$$

We evaluate the inverse of the distance between both frontiers at each point which results in  $n$  estimates of scale efficiency and compare each estimated scale efficiency with the appropriate  $p$ -values by means of bootstrapping (see algorithm, described in detail in Simar and Wilson, 1998 and 2002).

We define  $\hat{\omega}_{obs}$  as the observed value of the test statistic  $\hat{\omega}$  in a particular application. The  $p$ -value for  $H_0$  can be expressed by  $p = \Pr(\hat{\omega} \leq \hat{\omega}_{obs} \mid H_0)$ . Under  $H_0$  it is equivalent to  $p = \Pr(\hat{\omega} - \omega_0 \leq \hat{\omega}_{obs} - \omega_0 \mid H_0)$ . When we apply the bootstrap procedure the  $p$ -value can be approximated by  $\hat{p} = \Pr(\hat{\omega}^* - \hat{\omega} \leq \hat{\omega}_{obs} - \omega_0 \mid H_0)$  conditioned on the original sample. This also implies that  $\hat{\omega} = \hat{\omega}_{obs}$  and  $\hat{p} = \Pr(\hat{\omega}^* \leq 2\hat{\omega}_{obs} - \omega_0 \mid H_0)$ . Since  $\hat{\omega}$  is a consistent estimator, we obtain asymptotically  $\hat{p} = \Pr(\hat{\omega}^* \leq \hat{\omega}_{obs} \mid H_0)$ .<sup>10</sup> Therefore, we reject the null hypothesis for small  $p$ -values.

---

<sup>9</sup> For the Test 2, we only consider the distance to the NIRS frontier instead of CRS.

<sup>10</sup> Simar and Wilson (2002) show that the naïve bootstrap to construct pseudo samples yields inconsistent bootstrap estimation due to the estimation of boundaries of sets. The smooth bootstrap deals with this problem and gives consistent estimates. We use a univariate kernel estimator of density applying also the univariate reflection method of the original distance function estimates and then draw from this.

### 3 Data and Model Specification

#### 3.1 Data

Technical and physical data came from the utilities' annual reports as reported by the "Verband Deutscher Verkehrsunternehmen" (VDV). The VDV has approximately 440 members, where 360 public transit companies reported their data. Among the 360 public transport companies are 60 companies which are offering bus as well as regional rail services. We focus on "bus-only" companies that provide single output bus services;<sup>11</sup> therefore we deleted these companies from the sample. We created a consistent unbalanced panel data set for slightly more than the half of the available bus companies: It ranges from 127 to 179 German public transit bus companies in different years for the period 1990 to 2004 (see Table 2b for a summary of observations per year). The deletion from different companies was due to two ex ante data driven criteria and one ex post frontier driven criteria.

With regard to the ex ante data driven aspects we first of all deleted bus companies where at least one of our input-output variables were missing in the data set of the VDN. Second we deleted companies with an under proportionally low number of employees, due to the assumption of outsourced activities. Some data on outsourcing busses is available, but the statistics do not consider outsourcing of labor. We calculated the ratio of reported employees per busses. We assumed that companies with a ratio smaller than one (less than one employee per bus) outsource a high amount of services; therefore these companies were deleted from the sample.<sup>12</sup>

---

<sup>11</sup> Multi-output companies, also operating in metro or metropolitan railways are not considered in the framework of this analysis due to data availability; in particular, there is no precise information on how to allocate total employment to the different activities.

<sup>12</sup> In total, we deleted 365 observations during 1990-2004.

The frontier driven criteria consists in deleting companies from our sample by means of a super-efficiency analysis (Anderson and Peterson, 1993). Within this framework, decision-making utilities within the efficiency frontier might obtain an efficiency score greater than one because the bus firm itself cannot be used as a peer (see Coelli et al., 2005) and therefore cannot form part of its reference frontier. By means of this linear programming problem it is possible to identify the most efficient frontier firms and provide a ranking system. We use the super-efficiency approach to avoid the problem of sensitivity to extreme observations, common to all traditional nonparametric envelopment estimators of frontiers like the DEA. Extreme observation in the data set might inappropriately influence the estimation of the performance of other firms in the sample.<sup>13</sup>

In the empirical literature there exist other approaches to identify ex ante outliers and delete some extreme points from the sample before starting the estimation procedure (see e.g. Wilson, 1993, 1995; Simar, 2003). Here the problem consists in defining the number of outliers a priori.<sup>14</sup> We argue in favor of an ex post approach, the super-efficiency, with the focus on a well defined frontier, with a shape not driven by extremely efficient observations.

Within this framework we analyze the technical efficiency, focusing on physical inputs and outputs and the production processes. The German companies operate under different technical and institutional conditions. The service areas differ in customer density and the geographic circumstances. Therefore we include in our linear optimization problem a structural variable to capture the cost-disadvantages of firms operating in less favorable areas (see Section 3.2).

---

<sup>13</sup> We deleted firms from the sample which obtained under the approach of super-efficiency a score higher than 300%.

<sup>14</sup> Fox et al. (2004) and Ondrich and Ruggiero (2002) also derived sophisticated methods to detect outliers within the DEA framework. There are also other approaches in the empirical literature which focus on robust estimators that have been recently developed as alternatives to the traditional DEA frontiers (see Daraio and Simar, 2005, Aragon et al., 2002) They all use the concept of the so called concept of “partial” frontier (order-m, or order alpha quantile frontier), as opposed to the traditional idea of a “full” frontier that envelops all the data (see Simar and Wilson, 2007).

Network size, in terms of km line length, varies greatly. There are also considerable differences in consumer density (number of customers per km network length). Partial productivity indicators vary somewhat among the companies. The average labor productivity of all firms has increased during the observation period from 383 million passenger-km per employee in 1990 to 426 million passenger-km per employee in 2004. The individual firms themselves feature quite different labor productivity levels, e.g. from 1813 million passenger-km per employee (Aurich) to only 11 million p-km per employee (Uetersen) in 1990.<sup>15</sup>

### **3.2 Model specification**

Although the production process of bus services is complex, efficiency analysis is generally performed based on a simplified representation limited to one output and two inputs (labor and capital), and appended by structural parameters (for a survey of the main models used, see Brons et al., 2005 and De Borger et al., 2002). Our base model is also limited: one output (seat kilometers or bus kilometers) and two inputs: labor (number of full-time and part-time workers) and capital (number of busses). The definition of seat and bus kilometers is as follows: bus km is calculated by the number of busses times the network length times the frequency to circulate the network, seat km is defined as bus km times the number of seats. There exists also another frequent output measure: passenger km, which is defined as bus-km times passenger traveled in the bus. The use of passenger kilometers considers the demand side; seat kilometers as well as bus kilometers focus on the supply side of the production process. The present paper deals with the supply side (seat km or bus km). The capacity utilization does often not lie in the public transport company's area of influence. They are not

---

<sup>15</sup> This is partly due to the degree of outsourcing, in addition to what we accounted for in our ex ante outlier detection (see section above). Therefore we have to be careful in interpreting well performing companies, which might be due to outsourced activities.



directly responsible for advertisements, ticketing and traffic planning. So the focus of the analysis on the supply side is economically justified.

Urban bus transport is a standard example of a network industry. Recent literature focuses especially on the incorporation of network characteristics (see Basso and Jara-Diaz, 2003, 2005, 2006). Within a cost function framework the multidimensionality of output is accounted for and it is suggested that returns to scale estimation is crucially dependent on assumptions with respect to (and the characteristics of) the network structure. The transport industry structure can then be analyzed using different indices: economies of density and economies of scale with variable network size (see Basso and Jara Diaz, 2006). Detailed data on characteristics of the network structure are not available for the German urban bus transport; therefore we treat output as one dimensional within the efficiency analysis. All results and conclusion are drawn with respect to this one dimensional analytical framework.

However, in a subsequent model, we add a density index (number of inhabitants per line length of the company) to account for the structural differences and cost disadvantages of firms operating in less densely settled areas.<sup>16</sup> When taken as an additional input, the density index (DI) favors the efficiency scores of less densely inhabited regions. DI is defined as a non-discretionary input.

Table 1 summarizes the three models, i.e. variants of the base model with different output measures (Model 1: seat km, Model 2: bus km), and one extended model (Model 3: including the density index).

Table 1: Different model specifications

---

<sup>16</sup> This facilitates the comparison between large urban providers with densely settled areas and therefore a very favorable area and the bus companies from the rural areas with long distances but less densely settled areas. The structural variable is measured by the number of inhabitants divided by the line length.

We start with the application of the extended DEA-approach: the calculation of technical efficiency scores under the assumption of super-efficiency, because we wanted to delete ex ante extreme observations, thus delete outliers from the sample.

For all model specifications we first calculated the technical efficiency scores under CRS and VRS to detect in descriptive framework scale inefficiency. The CRS approach assumes that the size of the companies is flexible and that utilities are able to improve productivity not only by increasing technical efficiency but also by exploiting scale economies. Next, we statistically test the hypothesis if the technology features constant returns to scale by means of bootstrapping to interpret the empirical results. We also report the statistical inference concerning the individual efficiency scores estimating the bias and confidence intervals (again Simar and Wilson's bootstrap algorithm, 1998).

## **4 Results and Interpretation**

This section examines the results of our base model using seat km as the output (Model 1). Our major interest lies in the inference regarding the returns to scale characteristics of our sample in addition to our “back-of-the-envelope” analysis of increasing and decreasing returns to scale. We also comment briefly upon Model 2 using bus kilometers as output. We further examine the influence of different operating area characteristics, the density index (DI) (Model 3).

### **4.1 Results from the base model**

#### **4.1.1 General trends - VRS and CRS estimations**

We start with DEA Model 1 with capital and labor as input and seat km as output. We estimate an intertemporal frontier, more precisely a cross section pooled frontier, where each

observation is accounted for as a single company without considering any panel structure of the data. The summary statistics of the different model specifications are outlined in Table 2. For DEA Model 1 we obtain an average technical efficiency of 39.5% under the strict CRS assumption where we assume one optimal firm size. The VRS scores only represent the pure technical inefficiency of the bus companies. Therefore, we eliminate the scale effect and compare only companies within similar sizes. We now see that the companies gain in efficiency and we can obtain on average a score of 42.8% under VRS. We notice that the average efficiency is relatively low under both assumptions. The summary statistics are also given for each year 1990-2004 within a pooled intertemporal estimation. We notice a slight increase in technical efficiency over the years on average (42.0% - 46.0%).

Table 2: Summary statistics of results from different model specifications

Figure 1 shows the difference between the VRS and CRS scores over the pooled data sorted by size, from left (high output, measured in the base model by seat km) to right (low output). We see a clear trend: the greatest portion of the smaller companies shows significant differences between the two scores, an indication of scale inefficiencies. We calculated the average of the VRS-CRS difference for the 50% largest companies, 0.56%; and the difference for the 50% of the smallest, 6.06%. This clearly indicates scale inefficiencies of the smaller firms. We can derive one early conclusion from these results: smaller bus companies have a scale disadvantage.

Figure 1: Difference in results (VRS-CRS) ordered by size (seat km)

#### 4.1.2 Robustness of efficiency estimates

For sensitivity analysis we now apply the bootstrap algorithm to obtain confidence intervals and bias corrected efficiency scores. We verify that  $1/3(\hat{Bias}^2)/\hat{Variance}$  is well above unity, hence the bias correction can be used (see Simar and Wilson, 2000).

The bias estimate is defined as  $\hat{Bias}(\hat{\theta}_{VRS}) = B^{-1} \sum_{b=1}^B \hat{\theta}_{VRS}^* - \hat{\theta}_{VRS}$  with  $B$  as the number of replication,  $\hat{\theta}_{VRS}^*$  the bootstrapping values of the original estimate  $\hat{\theta}_{VRS}$ . The confidence intervals are determined by  $\Pr(\hat{c}_{\alpha/2} \leq \hat{\theta}_{VRS}^* - \hat{\theta}_{VRS} \leq \hat{c}_{1-\alpha/2} | \hat{P}) = 1 - \alpha$  with  $\hat{c}_{\alpha/2}, \hat{c}_{1-\alpha/2}$  defined as the estimated upper and lower quantiles.

After 2,000 replications we estimate a mean bias of 0.17, an estimated mean variance of 0.01. The bias and mean variance of the estimates are quite low after 2,000 replications, so we consider our results to be generally robust. Since we only wish to reflect general trends within the German bus sector, we do not consider in detail the individual bias of the individual efficiency estimates. We notice, however, a high degree of difference in the largest and smallest companies between the VRS estimates and the bias corrected results. This statistical inference reveals the need for careful interpretation of the optimal firm size. Since we are testing the hypothesis that “big is beautiful”, a more detailed analysis is required for the returns to scale characteristics as well.

#### 4.1.3 Tests on returns to scale

The bootstrap approach is also valuable for testing the returns to scale characteristics of our sample because it provides a statistical indication of which estimator gives more reliable results about the nature of the production technology and the individual efficiency scores. To make inferences about empirical applications, the asymptotic sample distributions of the envelopment estimators are required (see Simar and Wilson, 2000, 2007). In fact, the

bootstrap” algorithm remains the only practical way of making inferences when using the multivariate DEA approach. We therefore apply the first test  $H_0$  to find whether the returns to scale are constant and firms operate under optimal size (Test 1). Next, we test  $H_0$  to find whether the firms feature non increasing returns to scale (Test 2). We consider the entire pooled sample. We test Model 1 including two inputs (labor, number of buses) and one output (seat kilometers). The number of replications was 2,000.

In Test 1, we find a very low p-value of almost 0.001 for global constant returns to scale, indicating that there is no risk of falsely rejecting the  $H_0$ . In other words, scale inefficiency appears to be present in the German public transit sector and the majority of companies are not operating at an optimal scale. This situation requires us to test for individual constant returns to scale at the level of each firm. Figure 2 shows the p-value for each bus company, once again ordered by size (seat km) for the pooled sample from 1990 to 2004. The larger bus companies feature quite high p-values, thus to reject the  $H_0$  of individual constant returns to scale would imply a high type-one error. The smaller bus companies on the contrary, have much smaller p-values; so that the  $H_0$  can be rejected (with only a few exceptions). Some irregularities are observed in the middle range of the scale; this involve that also medium sized busses might operate under CRS.

We use Test 2 to discover if the production frontier is globally and individually NIRS ( $H_0$ ) or VRS ( $H_1$ ). After 2,000 replications we obtain a p-value of almost 0.000, and can reject the  $H_0$  and conclude that the underlying technology does globally not exhibit non-increasing returns to scale. Testing for individual NIRS, we find that larger bus companies feature high p-values which indicate NIRS (thus constant or decreasing). The p-values of the smaller bus companies are low, tending to zero. These results lead us to conclude that small bus companies are characterized by non constant, increasing returns to scale and that they could increase their

efficiency by adapting their size. In general, the bootstrap test confirms that technical efficiency increases with optimal firm size, e.g. extending - operation areas via merging. We reject the idea that companies only differ by their pure technical efficiency; size does matter.

Figure 2: Test for individual constant returns to scale

## **4.2 Model variations**

### **4.2.1 Different output definition (Model 2)**

There is an ongoing debate about the appropriate output measure for bus companies' efficiency. The -“technical- or physical” side favors seat km or bus km as the relevant outputs, in contrast to the service-oriented side which favors the utility of the served passengers, or passenger km. Pure supply indicators such as bus km or seat km reflect a technical output, but do not consider the services delivered, and thus the efficiency score of the respective bus company may be produced by running an empty bus for another tour. Stated simply, an “efficient” bus company may be one that does not serve a single passenger. On the other hand, service-oriented indicators such as passenger km may be misleading because they cannot be controlled for by the bus company. Berechman (1993) and De Borger et al. (2002) provide a more in-depth discussion of the issue.

In our case, we focus only on the supply indicators such as bus or seat kilometers. Therefore, in addition we ran the above model specifications using bus km (Model 2) as output. In Model 2 the average technical efficiency is 20.5% under the VRS assumption. We do notice a decline in the technical efficiency assuming bus km as an output variable. This might be explained by the fact that some extremely efficient companies define the frontier and that the difference to the frontier is therefore large for the majority of companies. But as we have two different linear optimization and thus two different frontiers we are not able to compare directly the average efficiency level. We are more interested in the relative ranking of

observations in both models. Optical inspection of the ranking of the companies as well as the statistical returns to scale test confirms the similarities of the results. We conclude that the scale inefficiency of the smaller companies is robust throughout different model specifications.

#### **4.2.2 Structural characteristics (Model 3)**

The above DEA-results may be “unfair” if they do not consider structural characteristics of individual companies, such as the density of the area served. In Model 3, we take into account the different operating characteristics of the bus companies including the density index (defined as the ratio of inhabitants per km of network length). Our sample contains large urban providers and companies operating in less favorable areas (characterized by a less densely settled rural landscape). We define density as an input, thus favoring the efficiency score of rural operating areas, which obtain “compensation”. Density is specified as a non-discretionary input because it is outside the control of the bus companies.

The average efficiency score (VRS) increases in Model 3 (45.5% under VRS). The slight increase of average efficiency is due to the additional constraint, adding one further input variable to the linear programming problem. This can be explained by the fact that increasing the number of dimensions will result in more observations lying on the boundaries of the DEA estimators, the efficiency frontier (see Simar and Wilson, 2007). But however, we can analyze the distributional effects of this model specification by comparing individual efficiency scores: Figure 3 shows the differences for DEA-VRS-estimations of Model 3 minus the DEA-VRS-estimations of Model 1, i.e. without the non-discretionary structural variable. The companies are ordered by the density index, starting with urban providers on the

left and ending up with small rural providers on the right. We can clearly see that smaller companies operating in rural areas benefit from the integration of the structural variable.<sup>17</sup>

Figure 3: Technical efficiency difference of Model 3 and Model 1

## 5 Conclusions

Efficiency analysis is an important instrument because it provides information to transportation companies and policy makers alike. It is also useful to reduce the information asymmetry between regulator and transport companies, and thus supports the policy process in this important sector. This paper applied nonparametric methods to analyze the technical efficiency of German public bus transport companies, using traditional approaches, such as simple CRS- and VRS-DEA, and advanced approaches for statistical inference techniques such as bias correction and returns-to-scale tests using bootstrapping mechanisms. We find increasing returns to scale for small- and medium-sized companies. This result is independent of the model formulation, and it also holds across different output variables (seat km and bus km). The introduction of a density index has a significant effect on the efficiency scores.

Our findings imply that the structure of the German public bus sectors may be improved by exploiting the scale economies. Small- and medium-sized companies clearly are at a disadvantage and should seek synergy effects primarily with neighboring companies. This may involve an optimization of the supply chain, e.g. joint sourcing of buses, but opportunities exist to exploit other cost-reducing measures such as sharing repair garages, joint purchase of pollution control equipment to modify existing fleets. Last but not least,

---

<sup>17</sup> We conducted the analysis in a descriptive way. Simar and Wilson (2001) provide a testing procedure for testing restrictions in frontier models by means of bootstrapping. They discuss statistical procedures for testing various restrictions for whether inputs or outputs are irrelevant, in addition they formulate tests of whether inputs or outputs may be aggregated. This is in particular important for possible dimension reduction when small data samples are available. As we dispose of a large data set we did not test the further restrictions to include the structural variable.



mergers between public bus companies should not be excluded as a policy option, even though they are more difficult institutionally.

## **6 References**

- Andersen, P., Petersen, N., 1993. A procedure for ranking efficient units in data envelopment analysis. *Management Science* 39, 1261-1264.
- Aragon, Y., Daouia, A., Thomas-Agnan, C., 2005. Nonparametric frontier estimation: a conditional quantile based approach. *Econometric Theory* 21 (02), 358-389.
- Banker, R.D, Morey, R.C., 1986. The use of categorical variables in data envelopment analysis. *Management Science* 32, 1613-1627.
- Basso, L.J., Jara-Díaz, S.R., 2003. Transport cost functions, network expansion and economies of scope. *Transportation Research E* 39, 271-288.
- Basso, L.J., Jara-Díaz, S.R., 2005. Calculation of economies of spatial scope from transport cost functions with aggregate output (with an application to the airline industry). *Journal of Transport Economics and Policy* 39(1), 25-52.
- Basso, L.J., Jara-Díaz, S.R., 2006. Is returns to scale with variable network size adequate for transport industry structure analysis? *Transportation Science* 40, 259-268.
- Berechman, J., 1993. *Public Transit Economics and Deregulation Policy*. North- Holland.
- Bjurek, H., Hjalmarsson, L., Førsund, F.R., 1990. Deterministic parametric and nonparametric estimation in service production. *Journal of Econometrics* 46, 213–227.
- Boame, A. K., 2004. The technical efficiency of Canadian urban transit systems. *Transportation Research E* 40(5), 401-416.
- Brons, M., Nijkamp, P., Pels, E., Rietveld, P., 2005. Efficiency of urban public transit: a meta analysis. *Transportation* 32, 1-21.

- Cazals C., Florens, J.P., Simar, L., 2002. Nonparametric frontier estimation: a robust approach. *Journal of Econometrics* 106, 1-25.
- Coelli, T., 2000. On the econometric estimation of the distance function representation of a production technology. Discussion Paper 2000/42, Center for Operations Research and Econometrics, Universite Catholique de Louvain.
- Coelli, T., Prasada Rao, D. S., Battese, G. E., 2005. *An Introduction to Efficiency and Productivity analysis*. 2<sup>nd</sup> Edition, New York, Springer.
- Cooper, W., Seiford, L. M., Zhu, J., 2004. *Handbook on Data Envelopment Analysis*, Boston, Kluwer Academic Publishers.
- Cowie, J., Asenova, D., 1999. Organization form, scale effects and efficiency in the British bus industry. *Transportation* 26, 231-248.
- Cowie, J., 2002. Acquisition, efficiency and scale economies – an analysis of the British bus industry. *Transport Reviews* 22(2), 147-157.
- Cullinane, K., Wang T.-F.-F., Song, D-W., Ji, P., 2006. The technical efficiency of container ports: comparing data envelopment analysis and stochastic frontier analysis. *Transportation Research A* 40, 354-374.
- Daraio, C., Simar, L., 2005. Conditional nonparametric frontier models for convex and non convex technologies: a unifying approach. LEM Papers Series 2005/12, Laboratory of Economics and Management (LEM), Pisa, Italy.
- De Borger, B., Kerstens, K., Costa, A., 2002. Public transit performance: What does one learn from frontier studies? *Transport Reviews* 22(1), 1-38.
- Deprins, D., Simar L., Tulkens, H., 1984. Measuring labor efficiency in post offices. In Marchand, M., Pestieau, P., Tulkens, H., (Eds.), *The Performance of Public Enterprises: Concepts and Measurements*. Amsterdam, North Holland: 243-267. Elsevier Science Publishing.

- Färe, R., Grosskopf, S., Lovell, C., 1985. (Eds.) *The Measurement of Efficiency of Production*, Boston: Kluwer-Nijhoff Publishing.
- Farsi, M., Filippini, M., Kuenzle, M., 2006. Cost efficiency in regional bus companies. in *Journal of Transport Economics and Policy* 40, 95–118.
- Ferrier, G., Lovell, C.A.K., 1990. Measuring cost efficiency in banking: econometric and linear programming evidence. *Journal of Econometrics* 46, 229-45.
- Filippini, M., Prioni, P., 2003. The influence of ownership on the cost of bus service provision in Switzerland: an empirical illustration. *Applied Economics* 35(6), 683-90.
- Fox, K.J., Hill R.J., Diewert W.E., 2004. Identifying outliers in multi-output models. *Journal of Productivity Analysis* 22, 73-94.
- Graham, D. J., Couto, A., Adeney, W.E., Glaister, S., 2003. Economies of scale and density in urban rail transport: effects on productivity. *Transportation Research E* 39(6), 443-458.
- Hanusch, H., Cantner, U., 1991. Produktion öffentlicher Leistungen: Effizienz und technischer Fortschritt, *Jahrbucher für Nationalökonomie und Statistik* 208, 369-384.
- Leuthardt, H., 1986. Die Wirtschaftlichkeit von Kraftomnibussen im Stadtverkehr. *Der Nahverkehr* 03, 44-54.
- Leuthardt, H., 2005. Betriebskosten von Linienbussen im systematischen Vergleich. *Der Nahverkehr* 11, 20-25
- Lovell, C.A.K., 1993. Production frontiers and productive efficiency, in Fried, H.O., Lovell, C.A.K., Schmidt S. (Eds,) *The Measurement of Productive Efficiency: Techniques and Applications*, Oxford University Press, New York, pp. 3-67.
- Ondrich, J., Ruggiero, J., 2002. Outlier detection in data envelopment analysis: an analysis of Jackknifing. *Journal of the Operational Research Society* 53, 342-346.
- Piacenza, M., 2001. Productive structure cost efficiency and incentives in the local public transport: A survey of theoretical and empirical issues. Ceris-CNR / HERMES research project I trasporti locali: dal sussidio all'incentivazione e all'efficienza.

- Simar, L., 2003. Detecting outliers in frontier model: a simple approach. *Journal of Productivity Analysis* 20, 391-424.
- Simar, L., Wilson, P., 1998. Sensitivity analysis of efficiency scores. How to bootstrap in nonparametric frontier models. *Management Science* 44, 49-61.
- Simar, L., Wilson, P., 2000. A general methodology for bootstrapping in nonparametric frontier models. *Journal of Applied Statistics* 27, 779-802.
- Simar, L., Wilson, P., 2001. Testing restrictions in nonparametric frontier models. *Communications in Statistics: Simulation and Computation*. 30, 161-186.
- Simar, L., Wilson, P., 2002. Nonparametric tests of returns to scale. *European Journal of Operational Research* 139(1), 115-132.
- Simar, L., Wilson, P., 2007. Statistical inference in nonparametric frontier models: Recent Developments and Perspectives. In: Fried, H., Lovell, K., Schmidt, S. (Eds.), *The Measurement of Productive Efficiency*, Second Edition, Oxford University Press.
- Wilson, P. W., 1993. Detecting outliers in deterministic nonparametric frontier models with multiple outputs. *Journal of Business & Economic Statistics* 11(3), 319-323.
- Wilson, P. W., 1995. Detecting influential observations in data envelopment analysis. *Journal of Productivity Analysis* 6, 27-45.
- Viton, P. A., 1981. A translog cost function for urban bus transit. *The Journal of Industrial Economics* 29(3), 287-304.

## Appendix

Table 1: Different model specifications

	Input Labor	Input Busses	Input Density Index	Output Seat km	Output Bus km
Model 1	I	I		I	
Model 2	I	I			I
Model 3	I	I	I (non-dis)	I	

Table 2: Summary statistics of estimation results

2a) Pooled sample

<b>Model</b>	<b>Number of Observation</b>	<b>Mean</b>	<b>Std. Dev.</b>
<b>Model 1 CRS</b>	2292	0.395	0.124
<b>Model 1 VRS</b>	2292	0.428	0.137
<b>Model 2 VRS</b>	2292	0.204	0.134
<b>Model 3 VRS</b>	2292	0.455	0.152

2b) Pooled sample sorted by year

<b>Year</b>	<b>Number of Observation</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>95% Conf. Interval</b>	
<b>1990</b>	128	0.420	0.013	0.394	0.446
<b>1991</b>	131	0.419	0.013	0.392	0.445
<b>1992</b>	156	0.405	0.011	0.384	0.426
<b>1993</b>	175	0.401	0.009	0.383	0.420
<b>1994</b>	179	0.412	0.009	0.394	0.430
<b>1995</b>	170	0.421	0.010	0.401	0.440
<b>1996</b>	173	0.430	0.010	0.410	0.450
<b>1997</b>	165	0.418	0.009	0.399	0.436
<b>1998</b>	162	0.433	0.010	0.413	0.453
<b>1999</b>	157	0.433	0.011	0.412	0.454
<b>2000</b>	153	0.433	0.011	0.411	0.455
<b>2001</b>	153	0.450	0.011	0.427	0.472
<b>2002</b>	130	0.448	0.013	0.422	0.474
<b>2003</b>	127	0.453	0.013	0.427	0.479
<b>2004</b>	133	0.460	0.013	0.435	0.485

Figure 1: Difference in results (VRS-CRS) ordered by size (seat km)

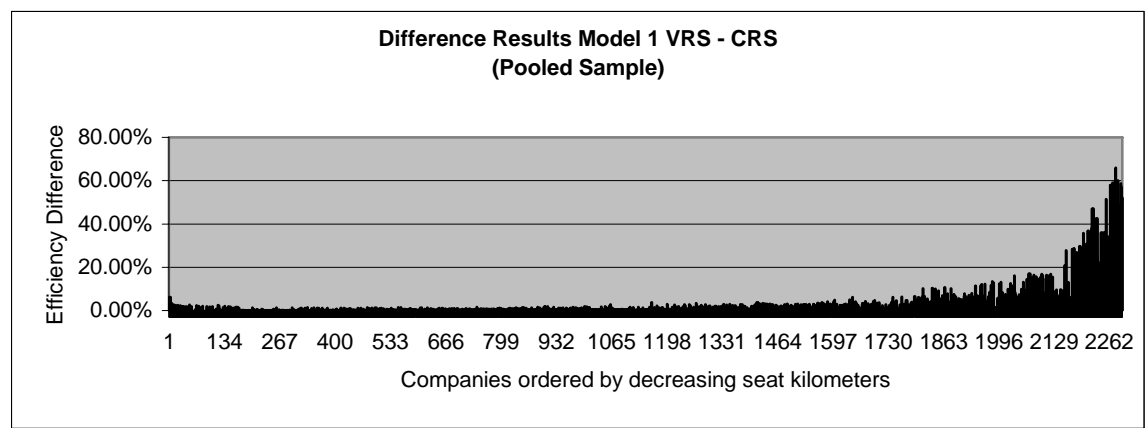


Figure 2: Test for individual constant returns to scale

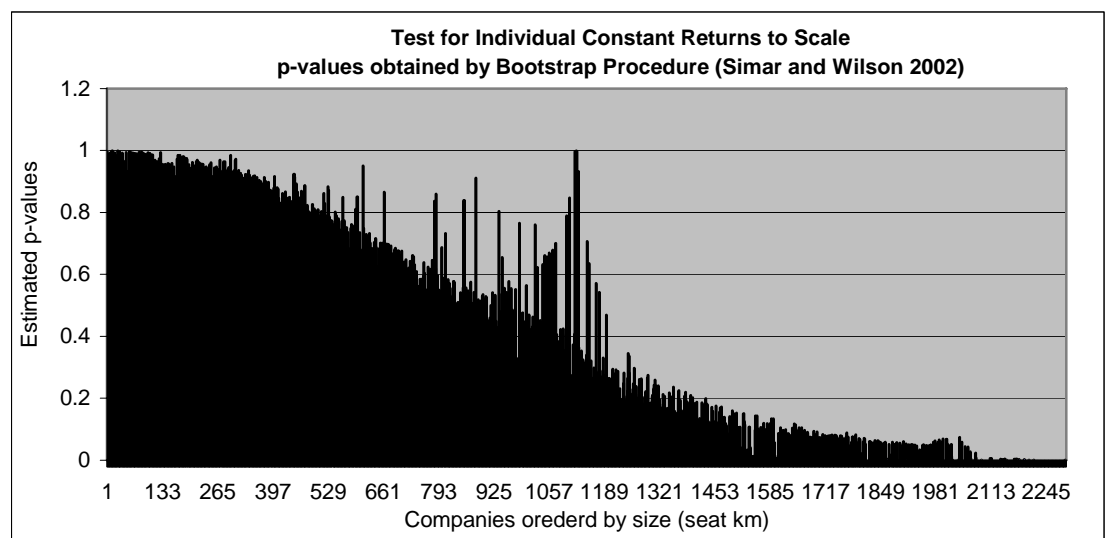
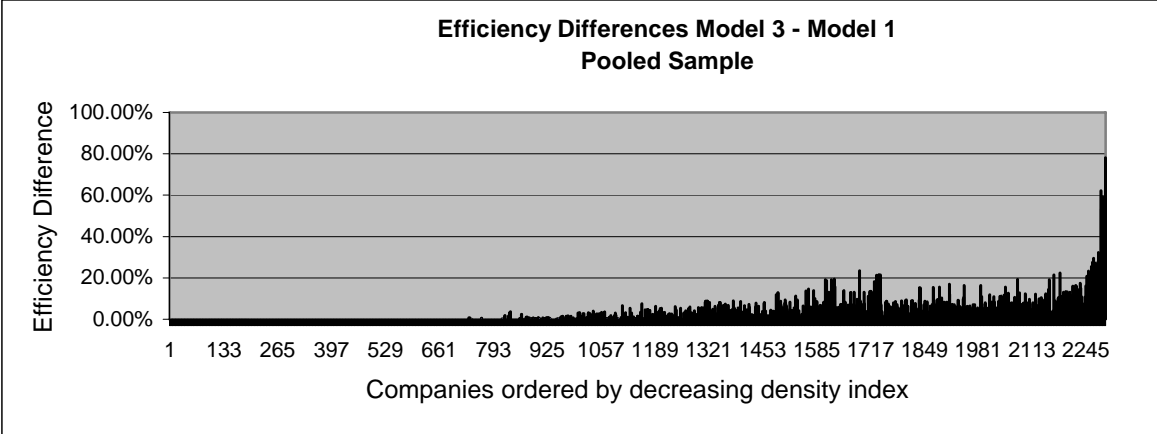




Figure 3: Technical efficiency difference of Model 3 and Model 1



## Acknowledgements

Christian von Hirschhausen is Professor of Energy Economics and Public Sector Management at Dresden University of Technology, and Research Professor at DIW Berlin (German Institute for Economic Research); Astrid Cullmann is Research Associate at DIW Berlin and Fellow of the DIW Graduate Program on Social and Economic Research. We thank Oleg Badunenko, Borge Hess, Gottfried Leuthardt, and Maria Nieswand for discussions and comments, and the Studienseminar “Efficiency” for support in data collection. The paper has been presented at the 5<sup>th</sup> Conference on Applied Infrastructure (INFRADAY) at Berlin University of Technology (October 2006), the 10<sup>th</sup> European Workshop on Productivity and Efficiency Analysis (EWEPA), July 2007 in Lille (France) and at the Annual Congress 2007 of Verein für Socialpolitik in Munich (Germany); the Annual Congress of the European Association for Research in Industrial Economics (EARIE) (2008). The usual disclaimer applies.